

Двухэтапное обучение СНС при использовании аугментации изображений для повторной идентификации человека

С.А.Игнатьева

Полоцкий государственный университет им. Евфросинии
Полоцкой
Полоцк, Беларусь
s.ignatieva@psu.by

Р.П.Богущ

Полоцкий государственный университет им. Евфросинии
Полоцкой
Полоцк, Беларусь
r.bogush@psu.by

Аннотация—Предложена технология обучения сверточных нейронных сетей (СНС) в приложении к задаче ре-идентификации людей по видеоданным распределенных систем наблюдения. Отличие заключается в том, что на первом этапе СНС предварительно обучается на аугментированной выборке изображений, а на втором этапе выполняется точная настройка, для которой пакеты изображений формируются без применения аугментации. Для корректировки весовых коэффициентов определяется величина ошибки каждой эпохи обучения. Показано, что такая технология обучения позволяет уменьшить значение функции потерь и повысить точность повторной идентификации

Ключевые слова — видеонаблюдение, ре-идентификация человека, предварительное обучение, точная настройка

1. ВВЕДЕНИЕ

В связи с тем, что на точность работы алгоритма ре-идентификации в значительной мере влияет качество обучения нейронной сети, используемой для извлечения отличительных признаков людей, необходима разработка эффективных подходов к обучению. Основной проблемой является недостаточное количество или разнообразие обучающих данных, что может приводить к снижению способности сверточной нейронной сети (СНС) к обобщению.

Поэтому актуальным является повышение точности алгоритмов ре-идентификации за счет разработки технологии обучения СНС. В настоящее время для увеличения качества обученной СНС, способной извлекать более эффективные отличительные признаки, при тренировке применяются перенос обучения (transfer learning) и аугментация данных.

Основная идея переноса обучения заключается в том, что на первом этапе для тренировки сети используется большой набор данных для схожей задачи. На втором этапе выполняется корректировка (точная настройка) весовых коэффициентов на изображениях из необходимого набора данных. Такой подход является достаточно эффективным, так как при обработке изображений с использованием СНС независимо от задачи в глубоких слоях извлекается информация, характерная для любых объектов: цвета, границы, линии, формы и т.д. Точная настройка сети позволяет скорректировать значимость различных признаков именно для ре-идентификации и улучшить точность работы алгоритма. Однако для этого требуется наличие большого размеченного набора данных и времени на

обучение. Известно, что предварительное обучение на ImageNet, состоящего из 3,2 миллиона размеченных образцов, может занимать несколько недель [1].

Аугментация увеличивает разнообразие набора данных и за счет этого при одном и том же составе выборки на разных эпохах обучения обрабатываются изображения с различными преобразованиями. Позволяет повысить точность работы СНС, однако значения функции потерь при обучении имеют более высокие показатели, чем при отсутствии аугментации. Это связано с тем, что видоизмененные изображения часто имеют нереалистичные преобразования, и не могут встречаться на реальных кадрах.

Для уменьшения значений функции потерь предлагается использовать технологию двухэтапного обучения, при которой на первом этапе используются аугментированные изображения из требуемой обучающей выборки, на втором – выполняется точная настройка (fine tuning) на исходных данных.

2. ТЕХНОЛОГИЯ ДВУХЭТАПНОГО ОБУЧЕНИЯ СНС

При обучении нейронной сети для задачи ре-идентификации используется пакетная обработка, при этом на вход СНС подается пакет из B изображений, обрабатываемых параллельно. Размер пакета обычно определяется в диапазоне от 16 до 64 изображений (X_i , Y_i), где X_i – матрица пикселей, Y_i – идентификатор человека.

На первом этапе пакеты формируются с применением метода аугментации из [2]. В каждом пакете могут присутствовать изображения, к которым применяется циклический сдвиг или преобразование в полутонное. Часть пакетов содержит изображения с фрагментами, замещенными уменьшенными копиями других изображений из пакета. При обработке каждого поступающего комплекта изображений вычисляется значение функции потерь L_{batch} по формуле (1), если в пакете содержатся замещенные фрагменты, и по формуле (2), если такое преобразование не применялось:

$$L_{batch} = E(X_a^{out}, Y_i) \cdot \lambda + E(X_a^{out}, Y_j) \cdot (1 - \lambda), \quad (1)$$

$$L_{batch} = E(X^{out}, Y_i), \quad (2)$$

где $E()$ – кросс-энтропийная функция потерь, X_a^{out} – изображение на выходе СНС с меткой Y_i , фрагмент которого замещен изображением с меткой Y_j ; X^{out} –

выходное изображение, для которого замещение фрагмента не осуществлялось.

Значения функции потерь L_{batch} для обрабатываемых пакетов суммируются, пока все данные обучающей выборки не пройдут через сеть, т.е. в течение одной эпохи обучения, для которой рассчитывается среднее значение функции потерь:

$$L_{epoch} = (\sum L_{batch} \cdot B) / S_{dataset}, \quad (3)$$

где $S_{dataset}$ – размер обучающей выборки.

После каждой эпохи выполняется корректировка весовых коэффициентов w классификационного слоя со скоростью обучения α :

$$w = w - \alpha \cdot (\partial L_{epoch} / \partial w), \quad (4)$$

где $\partial L_{epoch} / \partial w$ – градиент функции потерь. Весовые коэффициенты w_{hidden} скрытых слоев определяются выражением:

$$w_{hidden} = w_{hidden} - 0,1 \cdot \alpha \cdot (\partial L_{epoch} / \partial w). \quad (5)$$

Скорость обучения уменьшается после n -й эпохи. Предварительный этап обучения на аугментированных данных длится m эпох.

На втором этапе выполняется точная настройка весовых коэффициентов, для которой пакеты изображений (X_i, Y_i) формируются без применения аугментации. Вычисление значений функции потерь для пакета и эпохи вычисляется по формулам (2) и (3), корректировка весовых коэффициентов согласно выражениям (4) и (5) аналогично первому этапу. Такой подход позволяет уменьшить значение функции потерь и повысить точность повторной идентификации.

3. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Для оценки эффективности предлагаемой технологии двухэтапного обучения использовалась СНС ResNet-50 в алгоритме ре-идентификации [3]. Гиперпараметры при обучении: количество эпох $K_{epoch} = 80$, скорость обучения $\alpha = 0,07$, размер пакета $B = 16$, наборы данных для обучения: Market-1501, DukeMTMC-ReID, MSMT17. Точность оценивалась в метриках Rank1 и mAP. Для обучения алгоритма ре-идентификации рассматривались четыре подхода, результаты экспериментов для которых представлены в таблице 1. В таблице также приводятся значения функции потерь (loss) по окончании обучения.

Таблица 1. Влияние технологий обучения СНС на точность повторной идентификации человека

Технология обучения	Метрики	Набор данных		
		Market - 1501	Duke MTMC - ReID	MSMT 17
За один этап без аугментации	Rank1	0,8319	0,7285	0,4954
	mAP	0,6108	0,5253	0,2483
	loss	0,0393	0,0365	0,0477
С предобучением на ImageNet без аугментации	Rank1	0,8492	0,7468	0,5807
	mAP	0,6504	0,5442	0,2980
	loss	0,0375	0,0367	0,0433
С применением аугментации	Rank1	0,8616	0,7455	0,5824
	mAP	0,6780	0,5587	0,3124
	loss	0,0830	0,0661	0,0272
Предложенная	Rank1	0,8815	0,7590	0,6022
	mAP	0,7025	0,5823	0,3259
	loss	0,0397	0,0363	0,0429

Использование предварительно обученной модели позволяет повысить точность ре-идентификации по всем

метрикам, используемым для оценки на каждом исследуемом наборе данных (см. таблицу 1). Однако для многих архитектур нейронных сетей этот этап не реализован, а тренировка на ImageNet требует больших вычислительных ресурсов и времени на обучение. Отсутствием таких требований отличается подход к обучению с применением аугментации.

Обучение СНС на аугментированных данных так же позволяет повысить точность повторной идентификации в метриках Rank1 и mAP для трех исследуемых наборов данных. При этом, как видно из таблицы 1, значения функции потерь остаются слишком высокими, что говорит о том, что локальный минимум найден не оптимально.

Устранить этот недостаток позволяет применение технологии двухэтапного обучения, при которой аугментация применяется только на первом этапе. Точная настройка позволяет снизить значение функции потерь и повысить точность повторной идентификации людей как относительно модели без аугментации и предварительного обучения, так и относительно других исследуемых подходов.

4. ЗАКЛЮЧЕНИЕ

Эффективность предложенного двухэтапного обучения СНС на аугментированных данных подтверждается экспериментальными данными и позволяет увеличить точность как относительно базового алгоритма без аугментации, так и относительно модели, для обучения которой аугментация применялась на протяжении всех эпох. Кроме этого, как показали исследования, при использовании для обучения предварительно обученной на ImageNet модели показатели точности ниже, чем для предложенной технологии двухэтапного обучения. Преимуществом так же является отсутствие необходимости поиска предобученных моделей для редких или новых архитектур СНС. Двухэтапное обучение позволяет повысить точность повторной идентификации относительно базового алгоритма для Market-1501 в метрике Rank1 на 5,5 %, mAP на 15%; для набора данных DukeMTMC-ReID в метрике Rank1 на 4%, mAP на 11%. Наилучший результат наблюдается для набора данных MSMT17, для которого Rank1 увеличился на 21%; mAP на 31%.

ЛИТЕРАТУРА

- [1] Deng, J. ImageNet: A large-scale hierarchical image database / J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei // 2009 IEEE Conference on Computer Vision and Pattern Recognition. – 2009. – P. 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [2] Игнатова С.А. Аугментация данных для регуляризации в нейросетевых алгоритмах повторной идентификации людей по видеоданным / С.А. Игнатова, Р.П. Богуш, С.В. Абламейко // Материалы международного конгресса по информатике в трех частях, «Информационные системы и технологии» CSIST'2022. – 2022. – Т. 2. – С. 132 – 137.
- [3] Person ReID baseline pyTorch. . [Electronic resource]. — Access mode: https://github.com/layumi/Person_reID_baseline_pytorch (03.11.2022)